パネルデータを用いた配偶状態モデル

2010年7月9日　統計数理研究所　オープンハウス

井川　孝之　　総合研究大学院大学複合科学研究科統計科学専攻

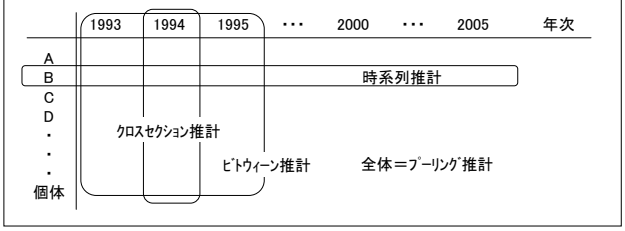
1. はじめに

我が国においては、依然として、非嫡出子の占める割合は2%弱と小さく、結婚を前提とした出産が社会規範となっている。また、近年の出生率低下の要因分解を行うと、有配偶率の低下によるところが大きい。ここでは、同一対象者を経年にわたり追跡調査したパネルデータを用い、有配偶状態がどのような要因によって説明されるかについて分析を行う。

2. パネルデータの特徴

パネルデータは、横断面データと時系列データ の双方の特性を併せ持ち、また、繰返し横断面データと比べても、同一個体の複数時点の状態を把握することができ、同一個体の変化を把握することが可能である。このように、複数個体の複数時点間の比較が可能な情報量が豊富なデータとなっていることがパネルデータの特徴である。

パネルデータの構造



海外のパネルデータとしては、現在まで継続している米国の家計に関する1964年に開始されたオハイオ州立大学と連邦政府統計局によるNational Longitudinal Survey（NLS）や1966年に開始されたミシガン大学のPanel Study of Income Dynamics（PSID）等があるが、日本のパネルデータは歴史も浅く数が少ない。

3. 今回使用するパネルデータ

今回、公益財団法人家計経済研究所が実施する「消費生活に関するパネル調査」（以下「家計研パネル」と言う）の個票データ（一部）の利用許可を得て、分析を行った。家計研パネルは、若年女性の生活実態を明らかとすることを目的として1993年に調査が始まった我が国のパネルデータのさきがけでもあり、現在まで継続している。

家計研パネルの対象者

満24～34歳の女性が対象。開始後5年毎に新しい標本が追加されている。

コーホートA	1993年度（第1年度）調査対象	1500人（開始時満24～34歳）
コーホートB	1997年度（第5年度）追加対象	500人（追加時満24～27歳）
コーホートC	2003年度（第11年度）追加対象	836人（追加時満24～29歳）

4. ロジット・プロビットによる要因分析

有配偶・無配偶の2値の状態変数を説明するモデルとして、まず、ロジット・プロビットモデルを取り上げる。被説明変数は、パネルデータの特徴を生かすため、翌年の配偶状態について検証を行うこととした。説明変数は、先行研究を参考に設定した。

$$y_{it}=\begin{cases}1 & (y^*_{it}>0)\\ 0 & (y^*_{it}\leq 0)\end{cases}$$

$$y^*_{it}=X'_{it}\cdot\beta+\mu_0+\nu_{it}$$

$$\Pr[y_{it}=1]=\Pr[y^*_{it}>0]=\Pr[\nu_{it}>-X'_{it}\cdot\beta-\mu_0]=F(X'_{it}\cdot\beta+\mu_0)$$

累積密度関数Fが標準正規分布の累積密度関数（Φ）であるものがプロビット・モデル、ロジスティック分布の累積密度関数（Λ）であるものがロジット・モデルである。

翌年の配偶状態を説明するプロビット・モデルの推定結果を下表に掲げる。学歴ダミー等の係数及び検定結果が年次によって異なる結果が得られた。ロジット・モデルについても同様の結果となった。

● プロビット・モデル 1994 年配偶状態～1993 年各説明変数

	推定値	漸近的ｔ値	
定数項	-3.7027	-7.537	***
市郡規模	0.4399	5.766	***
年齢	0.1349	8.270	***
大学・大学院卒	-0.2783	-1.826	.
短大・高専卒	-0.2031	-1.554	
専門・専修卒	-0.3290	-2.520	*
正規従業員	-0.6876	-4.960	***
パート	-0.5809	-3.630	***
嘱託	-0.7282	-2.750	**
年収ゼロダミー	0.8360	4.680	***
年収区分	-0.2088	-3.074	**
財産区分	-0.1291	-4.011	***
親居住関係	0.2187	9.819	***

有意水準：　'***' 0.001　'**' 0.01　'*' 0.05　'.' 0.1　' '

● プロビット・モデル 1998 年配偶状態～1997 年各説明変数

	推定値	漸近的ｔ値	
定数項	-3.6164	-9.892	***
市郡規模	0.2901	4.567	***
年齢	0.1101	11.080	***
大学・大学院卒	0.5531	2.928	**
短大・高専卒	0.6733	4.044	***
専門・専修卒	0.7224	4.449	***
正規従業員	-0.5367	-4.283	***
パート	-0.4028	-3.210	**
嘱託	-0.8180	-3.492	***
年収ゼロダミー	0.6659	4.589	***
年収区分	-0.2370	-4.283	***
財産区分	-0.1122	-4.464	***
親居住関係	0.1454	8.926	***

有意水準：　'***' 0.001　'**' 0.01　'*' 0.05　'.' 0.1　' '

● プロビット・モデル 2004 年配偶状態～2003 年各説明変数

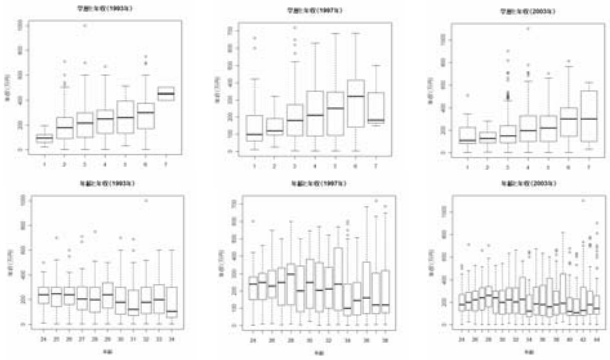
	推定値	漸近的ｔ値	
定数項	-3.0242	-10.448	***
市郡規模	0.1824	3.215	**
年齢	0.0816	12.780	***
大学・大学院卒	0.4138	2.327	*
短大・高専卒	0.5988	3.672	***
専門・専修卒	0.4468	2.788	**
正規従業員	-0.4125	-3.561	***
パート	-0.0011	-0.010	
嘱託	-0.6470	-4.013	***
年収ゼロダミー	1.0936	8.236	***
年収区分	-0.1993	-4.315	***
財産区分	-0.0721	-3.169	**
親居住関係	0.1236	8.710	***

有意水準：　'***' 0.001　'**' 0.01　'*' 0.05　'.' 0.1　' '

（注）親居住関係については、次のコードが付されている。
1　親と同一建物で、生計を共にしている（同居世帯）
2　親と同一建物で、生計が別（準同居世帯）
3　親と同一敷地内の別建物に居住（準同居世帯）
4　同一町丁内又は1km以内に親が居住（近隣地域居住）
5　（区のある13大都市居住者）同一区内に親が居住
6　（その他の市部、郡部居住者）同一市町村内に親が居住
7　同一都道府県内に親が居住
8　上記1～7地域以外に親が居住
9　親は全員死亡

5. 交互作用項についての検証

次に、前節で取り上げたロジット・プロビットにおいて、交互作用項を挿入することについて検討した。下図は、1993年・1997年・2003年の年収がゼロより大きい者を対象とした学歴と年収、年齢と年収の箱ひげ図であり、学歴が高いと年収が高い傾向が伺える。このことを踏まえ、学歴・所得を7区分とし、ロジット・プロビット分析を行った。



学歴と年収の交互作用項については、前節で取り上げたいずれの年次についても有意な結果とならなかった。

6. 混合効果モデルの適用

配偶状態を説明する要因には、個人毎の特性等が含まれる可能性がある。パネルデータは、分析方法を工夫することにより、個人毎の特性等を取り扱うことを可能とする。ここでは、配偶状態を説明する固定効果と変量効果の双方を取り扱うことが可能な混合効果モデル（Mixed Effect Model）の適用を試みた。

混合効果モデルの概要

$$y^*_i=\underset{\sim}{X}_i B+\underset{\sim}{Z}_i b_i+\varepsilon_i$$

固定効果　変量効果

$$X_i=\begin{pmatrix}x_{i11}&\cdots&x_{i1k}\\ \vdots&\ddots&\vdots\\ x_{im1}&\cdots&x_{imk}\end{pmatrix}\quad B=\begin{pmatrix}\beta_{11}&\cdots&\beta_{1k}\\ \vdots&\ddots&\vdots\\ \beta_{m1}&\cdots&\beta_{mk}\end{pmatrix}\quad y^*_i=\begin{pmatrix}y^*_{i1}\\ \vdots\\ y_{im}\end{pmatrix}$$

$$Z_i=\begin{pmatrix}z_{i11}&\cdots&z_{i1q}\\ \vdots&\ddots&\vdots\\ z_{im1}&\cdots&z_{imq}\end{pmatrix}\quad b_i=\begin{pmatrix}b_{i1}\\ \vdots\\ b_{iq}\end{pmatrix}$$

$$y^*_{ij}=\beta_{j0}+\beta_{j1}x_{ij1}+\ldots+\beta_{jk}x_{ijk}+\varepsilon_{ij}$$

$$\beta_{j0}=b_{i1}z_{ij1}+\ldots+b_{iq}z_{ijq}+\eta_j$$

次の3ケースについてモデルの推定及び比較を行った。AICを基準とすれば、プーリング推計ではなく、各対象者を各グループとみなして変量効果を扱った混合効果モデル（ケース1）が最も適当という結果となった。

	混合効果モデル等におけるｸﾞﾙｰﾋﾟﾝｸﾞ	AIC	残差の分散 σ^2_ε	変量効果の分散 σ^2_η
ケース１	i=1～5 j=1～873	1890	0.1636	1.2528
ケース２	i=1～5 j=1	2997	9.3369	0.0006
ケース３	ﾌﾟｰﾘﾝｸﾞﾌﾟﾛﾋﾞｯﾄ	2994	0.6774	－

対象年次：1993～1997 年（コーホート A）、継続回答数：873
固定効果は、ロジット・プロビットで取り上げた説明変数と同様のものを取り上げた。
ケース 1･2 においては、個人毎の変量効果を仮定。

7. 今後の課題

今回利用した家計研パネルは、我が国でも数少ないパネルデータの1つであるが、脱落サンプルに加え、調査年次によって非回答又は非該当等のデータが少なからず存在し、パネルデータの構造を利用したモデル策定の上で障害となることが分かった。非回答データ等にはバイアスが生じている可能性も否定できない。また、家計研パネルを利用したモデルによる分析結果を我が国全人口の有配偶状態の解釈へあてはめる場合は、標本に係る修正が必要であろう。パネルデータの特徴を生かした配偶・出生モデルの構築とともに、標本に係る修正が今後の課題である。